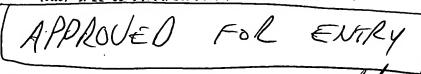
AN 10/613,261 Page 3



## AMENDMENTS TO AND LISTING OF CLAIMS

1. (Currently Amended) A method for generating training data that can be used with statistical models to normalize abbreviations in text, including:

providing a corpus of text including expansions of the abbreviations to be normalized; processing the corpus of text to identify the expansions in the corpus of text; processing the corpus of text to generate context information describing the context of the text in which the expansions were identified; and storing training data as a function of the generated context information, the training data including a set of feature vectors where each feature vector includes the abbreviation, associated expansion and the context information generated for

2. (Original) The method of claim 1 wherein: generating context information includes generating local context information; and storing training data includes storing local context data as training data.

the associated expansion identified in the corpus of text.

- 3. (Original) The method of claim 2 wherein the local context information and local context training data includes sentence level information.
- 4. (Original) The method of claim 3 wherein the sentence level information includes words in a sentence in which the identified expansion is located.
  - (Original) The method of claim 1 wherein:
    generating context information includes generating discourse context information;
    and
    storing training data includes storing discourse context data as training data.
- 6. (Original) The method of claim 5 wherein the discourse context information and discourse context training data include text section level information.

AN 10/613,261 Page 4

- 7. (Original) The method of claim 1 wherein:
- generating context information includes generating local context information and discourse context information; and
- storing training data includes storing local context data and discourse data as training data.
- 8. (Canceled)
- 9. (Currently Amended) The method of claim [[8]] 1 and further including processing text using a Maximum Er tropy model and the stored feature vectors to normalize abbreviations in the text.
  - 10. (Canceled)
- 11. (Original) The method of claim 1 and further including processing text using a statistical model and the stored training data to normalize abbreviations in the text.
  - 12. (Currently Amended) The method of claim 1 wherein:
  - the method further includes providing stored abbreviation data representative of abbreviations and associated expansions for which training data is to be generated; and
  - identifying processing the text to identify the expansions includes processing the corpus of text as a function of the stored abbreviation data.
  - 13. 15. (Canceled)